Florence, Italy 12-16 July 2015
ICSV22
The 22nd International Congress on Sound and Vibration

LASER SCANNING VIBROMETRY AND MODAL ANALYSIS TO CHARACTERIZE A VOCAL FOLD REPLICA

Paul Luizard, Nicolas Hermant, Xavier Laval, Xavier Pelorson
GIPSA-lab UMR 5216, 11 rue des mathématiques, BP 46, 38402 Saint-Martin-D'Hères, France
email: paul.luizard@gipsa-lab.fr

Fabrice Silva
Laboratoire de Mécanique et d'Acoustique, UPR CNRS 7051, 31 chemin Joseph Aiguier, 13402 Marseille cedex 20, France

Vocal folds are composed of elastic, soft, multilayer material, and are set to various vibration regimes during phonation, while speaking or singing. To explore such vibration phenomena, a vocal folds replica has been built, allowing to control physical parameters (subglottal pressure, vocal folds stiffness, and glottal aperture) in order to understand their respective contribution. Vocal folds are imitated by latex tubes filled with water under variable pressure. The present study aims at presenting mechanical measurements performed on a single vocal fold replica by means of a shaker provided with an accelerometer in conjunction with a laser vibrometer. This vibration measurement protocol yields a series of frequency response functions over a specific area of the vocal fold. Modal analysis is then performed using an algorithm based on the least square complex exponentials (LSCE) method, which has been developed for single input-multiple output (SIMO) systems. Results are further compared with those from the rational fraction polynomial (RFP) method. Although results are in fair accordance, the observed discrepancies are quantified and discussed.

1. Introduction

Voice production is of primary importance in the context of human communication. Although vocal folds mechanics is one of the elements involved in voice production [16], aspects of the dynamical behavior of this soft, elastic, multilayer material remain unclear. Vocal folds have been modeled by means of analytical models, e.g. the largely used two-mass model [9] and numerical models, e.g. using the finite element method [1]. Characterization of vocal folds is not trivial because of the vocal folds location, in the throat, not easily accessible by measurement devices. Previous research has focused on excised larynges whose mechanical properties change fast [5], especially the viscoelasticity, since those are early dead human tissue. An alternate means to perform measurements of vocal folds vibration in auto-oscillation regime is to use a replica [14]. Various geometries and materials have been used, based on the works from the musical acoustics community and the development of an artificial mouth [7]. Vocal Fold replicas allow for measuring frequency response functions (FRF) depending on various parameters such as the vocal fold geometry and stiffness.

Modal analysis applied to an element of human body has been seldom used in previous studies. Most research using this method is applied to bones quality or integrity [6, 15] and generally to hard material, as opposed to soft human tissue. Modal analysis has been applied to human vocal folds through extraction of empirical orthogonal functions [10], either based on numerical methods such as the finite element method [2] or based on high-speed glottography and videokymography [17] for in vivo experiments [12]. These studies have shown various particular vibration modes related to healthy and pathological vocal folds, including biphonation due to vocal fold asymmetries.

The present study focuses on vibration patterns of a single vocal foldreplica presented in section 2, in order to obtain a precise knowledge of the vibrational behavior of the vocal fold, independently from the rest of the replica, before conducting more complex experiments. The associated modal analysis methods used to characterize the modal behavior of the structure, both in time and frequency domains, are presented in section 3. Results from both methods are further discussed and compared in section 4.

2. Measurements

2.1 Vocal fold replica

The vocal fold replica under study is based on the design of previous research [14]. It is composed of a steel structure representing a half cylinder, covered by latex tubes filled with water. A schematic of these artificial vocal folds can be seen in Fig. 1a. Water inside the vocal folds is connected to a tank whose height controls the water pressure within the latex tubes, modifying its stiffness. These vocal folds are connected to a large box containing pressurized air representing the lungs. The whole replica allows for controlling independently three physical parameters: the vocal folds stiffness, upstream air pressure, and the distance between vocal folds.

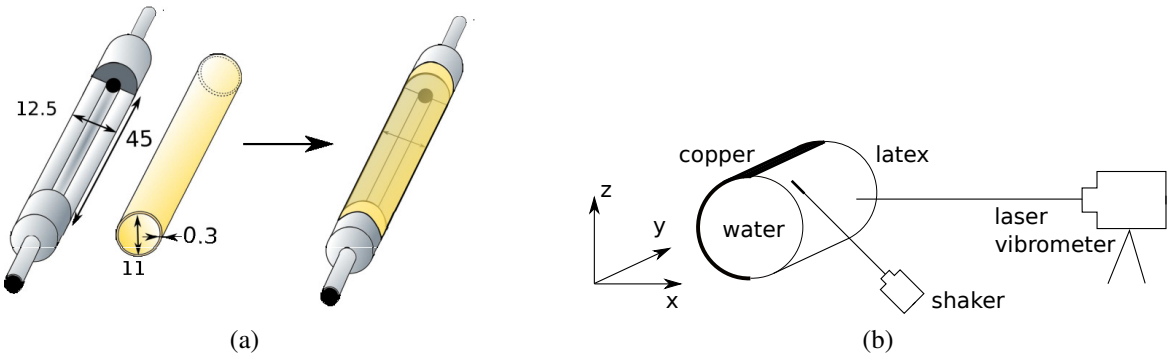

Figure 1: (a) Schematics of the vocal fold replica as a water-filled latex membrane on a steel structure (from [8], dimensions in mm), and (b) measurement setup with a tangent mechanical excitation and a normal vibrometry measurement.

2.2 Measurements procedure

Measurements are performed on a single vocal fold replica mounted on a half cylinder shell made of copper, as can be seen in Fig. 1b. The vibratory excitation is made by a shaker whose input signal is noise, ranging from 50 Hz to 400 Hz at a sampling frequency of 10 kHz. This noise was pre-filtered according to the power density spectrum of the shaker in order to provide a flat excitation spectrum. The shaker is in tangential contact with the latex tube near one end of the latter to leave some space to observation points, and it makes a 45° angle with both x and z axis, as shown in Fig. 1b. It should be noted that no attachment of the shaker rod is possible due to the soft and lightweight structure under study.

Observations are made with a scanning laser vibrometer allowing to perform velocity measurements with the same excitation signal at each point of a specific grid on the vibrating structure. An accelerometer is mounted on the shaker so that the excitation acceleration is known. In conjunction with the axial velocity of observation points measured by the vibrometer, transfer functions can be estimated using the ratio of the input and output cross spectral density over the input auto spectral density. This ratio provides a complex mechanical frequency response function (FRF) at each measurement point of the structure. The obtained FRF series, including amplitude and phase signals, as visible in Fig. 2, are used to estimate the modal parameters of the structure.

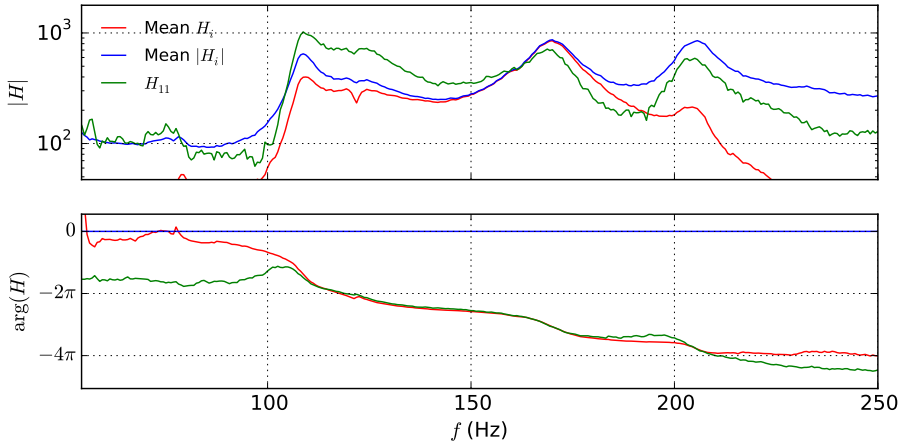

Figure 2: Example of a single FRF (green curves, on point 11), overall average of complex FRF (red curves) and average on FRF magnitudes (blue curves).

3. Modal analysis

Modal analysis used to characterize the vibratory behavior of a structure. Although vocal folds are naturally set to auto-oscillation by the air flowing between them, due to fluid-structure interaction and viscous forces, the present study considers a single vocal fold without air flow interaction. The natural resonances of this complex structure are studied by means of FRF measurements and modal analysis.

Two methods are used in the present study: the Least Square Complex Exponentials method (LSCE) and the Rational Fraction Polynomial method (RFP). The main difference between them is the fact that LSCE has been developed for time domain data while RFP uses frequency domain data [11]. In general terms, time domain models tend to provide the best results when a large number of modes exist in the data, whereas frequency domain models tend to provide the best results when the frequency range is limited and the number of modes is relatively small [11]. Both LSCE and RFP methods are said to be *indirect methods* because they are based on identification of the modal parameters of a model, as opposed to considering directly the general matrix equation of dynamic equilibrium. Besides, they allow for multiple degree of freedom (MDOF) analyses, i.e. they can detect several modes within a given FRF.

3.1 Time domain SIMO method

Least Square Complex Exponentials method (LSCE) method [4] is a single input-multiple output (SIMO) method, working in time-domain with one excitation point and processing simultaneously several FRF related to various measurement points. This method has been implemented in the Abravibe Toolbox (Matlab) [3], which was adapted to the requirements of the present study.

In the frequency domain, the FRF in terms of receptance α_{jk} (displacement at point j due to a force at point k) for a linear, viscously damped system with N degree-of-freedom can be written as

$$\alpha_{jk}(\omega) = \sum_{r=1}^{2N} \frac{{}_rA_{jk}}{\omega_r\xi_r + i(\omega - \omega_r\sqrt{1-\xi_r^2})}, \tag{1}$$

where ω_r is the natural frequency, ξ_r is the viscous damping factor, and ${}_rA_{jk}$ is the residue corresponding to each mode r. It should be noted that the sum order goes up to $2N$ because the N complex conjugate of the numerator are directly included since absolute value is considered. The corresponding impulse response is obtained by means of the inverse Fourier transform:

$$h(t) = \sum_{r=1}^{2N} A_r e^{s_r t}, \tag{2}$$

where $s_r = -\omega_r\xi_r + i\omega_r\sqrt{1-\xi_r^2}$. By discretizing the time length into L intervals Δt, the impulse response h becomes

$$h_n = h(n\Delta t) = \sum_{r=1}^{2N} A_r V_r^n, \tag{3}$$

where $V_r = e^{s_r\Delta t}$, which can be known using Prony's method [13]. The principle is that there always exists a polynomial in V_r with real coefficients β such that:

$$\beta_0 + \beta_1 V_r + ... + \beta_n V_r^n + ... + \beta_L V_r^L = 0. \tag{4}$$

In order to calculate the coefficients β to evaluate V_r, Eqs 3 and 4 yield, for each V_r:

$$\sum_{j=0}^{L} \beta_j h_j = 0. \tag{5}$$

L is taken as equal to $2N$ for convenience, so there are $2N$ sets of data points h_j, each one shifted one time interval, and β_{2N} is set to 1. This yields, in matrix form, for p impulse responses:

$$\mathbf{h}_m\ \beta = h'_m. \tag{6}$$

The least-squares solution can be found via the pseudo-inverse technique:

$$\beta = \left([\mathbf{h}_m]^T\mathbf{h}_m\right)^{-1}[\mathbf{h}_m]^T h'_m. \tag{7}$$

Knowing the coefficients β, a polynommial solver can be used to calculate the roots V_r from Eq. (4). The latter allows for using Eq. (3) to calculate the residues, and consequently the modal constants and phase angles.

The main issue of this method is the estimation of the number of modes within the series of FRF. Calculating the rank of matrix $\mathbf{h}_m$ from Eq. (6) can be used as an indication of that quantity.

3.2 Frequency domain SIMO method

A common basis of several frequency-domain methods is the modeling of the FRFs as the fraction of a numerator and a denominator that are expanded on basis functions Ω_n of the frequency ω (normalized by the upper bound of the frequency range under analysis):

$$\tilde{H}^k(j\omega) \sim \frac{Num^k(j\omega)}{Den(j\omega)} = \frac{\sum_{n=0}^{M} b_m^k \Omega_m(j\omega)}{\sum_{n=0}^{N} a_n \Omega_n(j\omega)} \tag{8}$$

with a common denominator Den. A least-squares formulation considers the error criterion

$$J = \sum_k \sum_i \left| \tilde{H}^k(\omega_i) Den(j\omega_i) - Num^k(j\omega_i) \right|^2 \tag{9}$$

that may be seen as weighting the approximation error by the denominator values. Denoting b^k and a the column vectors of the numerator and denominator coefficients, respectively, the minimization of the error criterion with respect to (b^k) leads to the first set of normal equations $Y \cdot b^k = X^k \cdot a$ with:

$$Y_{mn} = \Re(< \Omega_m, \Omega_n >) = \Re\left(\sum_i \Omega_m^*(j\omega_i) \Omega_n(j\omega_i) \right) \text{ and } X^k_{mn} = \Re\Big(< \Omega_m, H^k \Omega_n > \Big). \tag{10}$$

The inversion behaves bad when considering an expansion on the canonical polynomials, thus justifying the use of the Fourier functions $\Omega_n(j\omega) = e^{jn\omega}$ in the case of the Least-Squares Complex Frequency-domain method (LSCF, see [18]) or of the Forsythe orthogonal polynomials in the case of the Rational Fraction Polynomial method (RFP, see [19]). The error criterion then only depends on the denominator coefficients $J = a^T \cdot A \cdot a$ where

$$\Re(A) = \sum_k \left(Z^k - X^{kT} \cdot Y^{-1} \cdot X^k \right) \text{ with } Z^k_{mn} = \Re\Big(< H^k \Omega_m, H^k \Omega_n > \Big). \tag{11}$$

Contrary to Ref. [19], the same functions are used here for the numerators and the denominator, avoiding the expensive computation of several orthogonal polynomial basis and the formulation of a secondary least-squares approach to merge estimations from several FRFs. While the highest order coefficient is usually set to 1, Ref. [18] considers varying the denominator coefficient that is constrained to 1. In both cases, the minimization of J with respect to the remaining coefficient leads to the second set of linear normal equations using submatrices of A, and the poles, i.e. the roots of the denominator, are efficiently computed by means of the generalized companion matrix (see [20]) of the orthogonal polynomials in the case of the RFP, avoiding the numerical instability of the canonical polynomials.

It is interesting to note that the frequency-domain SIMO methods derived from Eq. 9 is known [21] to be biased for noisy measurements due to the weighting by the unknown denominator coefficient.

3.3 Application to measurements

Mobility measurements were performed on 40 points of the artificial vocal fold, as presented in Fig. 1b, providing 40 FRF. Both LSCE and RFP methods were used to analyze these transfer functions, providing complex pole values. The usual stabilization chart relies on N_{sim} evaluations by either increasing the model order N (in the RFP and the LSCE method) or varying the denominator coefficient a_n that is constrained to equal 1 (in the RFP method only). The classical technique appears to only retain a single estimation of a pole without really exploiting the various samples that could be identified as the same pole, at least as performed in Abravibe [3]. We here adopt a bi-dimensional dispersion representation in the (f_n, Q_n) plane, f_n being the resonance frequency and Q_n the quality factor. The stable poles whose real part is negative with positive frequency are presented in Fig. 3, where the distribution is heterogeneous.

Some regions show an important density of poles. One expects the agglomeration of samples to be related to physical poles, while the dispersion relates to non-physical ones. The LSCE graph shows a strong "background" noise, i.e. many samples that do not agglomerate. Moreover, the dense regions seem to have a larger extent in the Q_n axis, i.e. a higher dispersion of the evaluation of the quality factor. In contrast, the RFP method with variable constraint shows less "background" noise and more compact dense regions. This is in accordance with the observation of Ref. [18]

In order to quantify these results, a cluster analysis was performed to determine which estimated poles must be considered.

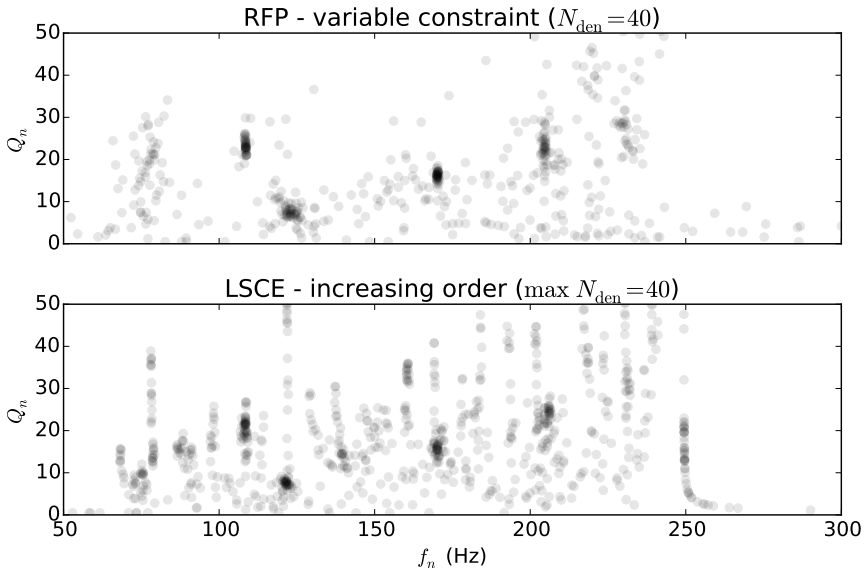

Figure 3: Representation of the stable poles (negative real part) with positive frequency, as the quality factor versus the resonance frequency, from both LSCE and RFP analysis methods, between 60 Hz and 250 Hz.

3.4 Cluster-based algorithm for modal analysis with uncertainties

In the manner of the procedure described in Ref. [18], we investigate the use of cluster algorithm to discriminate the physical poles from the nonphysical ones. Either in the s-plane or in the (frequency-quality factor) plane, physical poles gather into tight clusters while the others are scattered within a larger domain on the plane. The challenge is then to be able to identify the clusters and to compute statistics as theirs centroïd (the pole estimation) and standard deviation (the estimation uncertainties) along each axis. We do not follow the algorithm from Ref. [18], as it is explicit how nonphysical poles are gathered in clusters too.

Two approaches are proposed here. The first one relies on the Kernel Density Estimation (KDE) method. A histogram-like function Φ is built from the summation of contributions from each samples (f_i, Q_i) weighted by a gaussian window whose bandwidths are chosen to equal the expected resolution Δf and ΔQ on each axis:

$$\Phi(f,Q) = \sum_i \exp -\frac{1}{2}\left(\left(\frac{f-f_i}{\Delta f}\right)^2 + \left(\frac{Q-Q_i}{\Delta Q}\right)^2\right). \tag{12}$$

It is computed on a regular mesh of the domain of interest as defined by the excited frequency range ($50--400$ Hz) and pertinent quality factor (0-50). High values are found where samples agglomerate, and low values where samples scatters. All peaks higher than a threshold based on N_{sim} are then considered: it is required that at least 25% of the evaluations return a pole in the vicinity of the peak. They are iteratively fitted on a gaussian function whose moments (mean value and standard deviation on each axis) define the pole estimation and uncertainty (i.e. its confidence bounds).

The alternative method is based on a k-nearest-neighbours (k-NN) analysis. We look for the poles whose k nearest neighbours lie within the smallest spheres (using a quasi euclidian measure that accounts for the expected resolution in each axis). Typical values are searching for $k = N_{sim}/2$ neighbours. For each sphere, statistics on the distribution of the poles result in the final pole estimation and related uncertainty.

Table 1 compares estimations of the poles for RFP evaluations with a variable LS constraint for denominator order set to 40.

Table 1: Pole estimation for RFP and LSCE evaluations by means of KDE and k-NN methods.

		KDE method		k-NN method	
		Frequency (Hz)	Quality factor	Frequency (Hz)	Quality factor
RFP	Pole 1	108.5 ± 1.0	23.2 ± 2.7	108.5 ± 0.7	22.9 ± 2.4
	Pole 2	123.1 ± 2.4	7.5 ± 2.5	122.2 ± 1.6	7.4 ± 1.5
	Pole 3	170.2 ± 1.1	16.1 ± 2.3	170.2 ± 0.5	16.3 ± 1.3
	Pole 4	204.6 ± 1.2	22.8 ± 3.7	204.5 ± 1.5	22.8 ± 3.3
LSCE	Pole 1	108.4 ± 1.1	20.6 ± 3.4	108.3 ± 0.9	21.1 ± 2.6
	Pole 2	121.6 ± 1.4	7.4 ± 2.4	121.5 ± 1.7	7.5 ± 1.7
	Pole 3	170.2 ± 1.4	15.8 ± 2.9	170.0 ± 1.7	15.7 ± 2.2
	Pole 4	205.3 ± 2.0	23.3 ± 3.5	205.4 ± 2.7	24.0 ± 3.4

4. Conclusion

The vocal folds replica provides an interesting use-case of modal analysis. Modal density, consequent measurement noise and moderate to high damping allowed us to confront several methods to evaluate the poles given a model and parameters, such as for example the over-determined number of poles to look for. The results presented here reinforced our conviction that we can not rely on a single estimation of the poles, and that it is necessary to also estimate confidence bounds along with the modal parameters.

Two cluster approaches have been proposed in this paper. Both rely on the (f_n, Q_n) representation, and consider the multiple estimations of poles as samples of a multimodal distribution. The Kernel Density Estimation method is fully automated in the sense that the thresholds are automatically tuned to the density, the number of single estimations and the maximum denominator order considered. On the contrary, the k-nearest neighbors method still requires user interaction to select the optimal maximal distance and thus the maximal size of the cluster. When applying on results from either RFP or LSCE methods, the clustering methods show strong overlap which we interpret as a measure of quality of the two methods.

Moreover, in accordance with Ref.[18], we found that the frequency-domain method with variable constraint provides a clearer picture of the pole, as opposed to increasing order procedure, for time- or frequency-domain methods. However, after applying the cluster procedure, both methods lead to similar modal parameter evaluations and similar confidence bounds.

Acknowledgments

This work was supported by grant VoFoCam (ANR-12-PDOC-0018) of the ANR french agency.

REFERENCES

1. F. Alipour, D.A. Berry, and I.R. Titze. A finite-element model of vocal-fold vibration. *J. Acous. Soc. Am.*, **108** (6), 3003–3012, 2000.

2. D.A. Berry, H. Herzel, I.R. Titze, and K. Krischer. Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions. *J. Acous. Soc. Am.*, **95** (6), 3595–3604, 1994.

3. A. Brandt. *Noise and vibration analysis: Signal analysis and experimental procedures*. John Wiley and sons, Chichester, UK, 2011.

4. D.L. Brown, R.J. Allemang, R. Zimmerman, and M. Mergeay. Parameter estimation techniques for modal analysis. SAE Technical Paper Series, No. 790221, 1979.

5. R.W. Chan and I.R. Titze. Effect of postmortem changes and freezing on the viscoelastic properties of vocal fold tissues. *Ann. Biomedical Eng.*, **31**, 482–491, 2003.

6. B. Couteau, M.-C. Hobatho, R. Darmana, J.-C. Brignola, and J.-Y. Arlaud. Finite element modelling of the vibrational behaviour of the human femur using ct-based individualized geometrical and material properties. *J. Biomechanics*, **31**, 383–386, 1998.

7. J.S. Cullen, J. Gilbert, and M. Campbell. Brass instruments: Linear stability analysis and experiments with an artificial mouth. *Acta Acustica united with Acustica*, **86** (4), 704–724, 2000.

8. N. Hermant. *Observations, modélisation et simulation des vibrations des maquettes de plis vocaux. Applications à des configurations pathologiques.* PhD thesis, Grenoble INP, 2014.

9. K. Ishizaka and J.L. Flanagan. Synthesis of voice sounds from a two-mass model of the vocal cords. *Bell System Technical Journal*, **51**, 1233–1268, 1972.

10. E.N. Lorenz. Empirical orthogonal functions and statistical weather prediction. Technical report, Massachusetts Institute of Technology, 1956.

11. N. Maia and J. Silva. *Theoretical and experimental modal analysis*. Research Studies Press, Baldock, Hertfordshire, England, 1998.

12. J. Neubauer, P. Mergell, Eyscholdt, and H. Herzel. Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes. *J. Acous. Soc. Am.*, **110** (6), 3179–3192, 2001.

13. R. Prony. Essai expérimental et analitique sur les lois de la dilatabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, à différentes températures. *Journal de l'école Polytechnique*, **1** (2), 24–76, 1795.

14. N. Ruty, X. Pelorson, A. Van Hirtum, I. Lopez-Arteaga, and A. Hirschberg. An in vitro setup to test the relevance and the accuracy of low-order vocal folds models. *J. Acous. Soc. Am.*, **121** (1), 479–490, 2007.

15. W.R. Taylor, E. Roland, H. Ploeg, D. Hertig, R. Klabunde, M.D. Warner, M.C. Hobatho, L. Rakotomanana, and S.E. Clift. Determination of orthotropic bone elastic constants using FEA and modal analysis. *J. Biomechanics*, **35**, 767–773, 2002.

16. I.R. Titze. *Principles of Voice Production*. National Center for Voice and Speech, Iowa City, USA, 2000.

17. J.G. Švec and H.K. Schutte. Videokymography: High-speed line scanning of vocal fold vibration. *J. Voice*, **10** (2), 201–205, 1996.

18. Verboven, P., Cauberghe, B., Parloo, E., Vanlanduit, S. and Guillaume, P., User-assisting tools for a fast frequency-domain modal parameter estimation method, *Mech. Syst. Signal Proc.*, **18** (4), 759–780, 2004.

19. Richardson, M.H. and Formenti, D.L., Parameter estimation from frequency response measurements using rational fraction polynomials, *1st IMAC*, 167–186, 1982.

20. Day, D. and Romero, L., Roots of polynomials expressed in terms of orthogonal polynomials, *SIAM J. Num. Analysis*, **43** (5), 1969–1987, 2005.

21. Carcaterra, A. and Sestieri, A., Bias in modal parameters using the direct and iterative RFP identification procedures, *Meccanica*, **30** (1), 77–92, 1995.